Claims

We Claim:

1. A computer-implemented method for adaptively filtering URL messages routed across a network, the URL messages rejected based on a set of rules, the method comprising:

maintaining a frequency for each instance of a URL component, wherein the frequency is a function of a number of occurrences with which a URL component was rejected by a rule and a number of occurrences with which descendants of the URL component were rejected with the rule;

selecting a URL component according to a set of constraints; and

generating an exception rule for the selected URL component and its descendants.

2. The method of claim 1, wherein the set of constraints is selecting a URL component with a frequency exceeding a threshold and having no children with a frequency above the threshold.

3. The method of claim 1, wherein the set of constraints is selecting a URL component with a frequency exceeding a threshold.

4. The method of claim 1, further comprising applying the exception rule to determine whether to allow the selected URL component and its descendants.

5. The method of claim 2, wherein the threshold is a product of a total number of URL messages over a time interval and a percentage of the messages that should be allowed.

6. The method of claim 1, wherein the exception rule is generated by inferencing a scalar data type of the descendants of the selected URL component.

7. The method of claim 1, wherein the frequency is a direct count of the occurrences of the URL component.

8. The method of claim 1, wherein the frequency is a weighted count of the occurrences of the URL component.

9. A computer-implemented method for adaptively filtering URL messages routed across a network, the URL messages rejected based on a set of rules, the method comprising:

storing rejected URLs in a trie structure, wherein each node in the trie structure is associated with a URL component;

maintaining a frequency for each node associated with a URL component, wherein the frequency is a function of a number of occurrences with which a URL component associated with a node was rejected with a rule and a number of occurrences with which descendants of the URL component were rejected with the rule;

selecting a node in the trie structure according to a set of constraints; and

generating an exception rule for the selected node and its descendants.

10. The method of claim 9, further comprising applying the exception rule to determine whether to allow the selected node and its descendants.

11. The method of claim 9, wherein the set of constraints is selecting a node with a number of occurrences exceeding a threshold.

12. The method of claim 9, wherein the set of constraints is selecting a node with a number of occurrences exceeding a threshold and having no children with a number of occurrences above the threshold.

13. The method of claim 11, wherein the threshold is a product of a total number of URL messages over a time interval and a percentage of the messages that should be allowed to pass.

14. The method of claim 9, wherein the exception rule is generated by inferencing a scalar data type of the descendants of the selected URL component.

15. The method of claim 9, wherein the frequency is a direct count of a number of occurrences of the URL component.

16. The method of claim 9, wherein the frequency is a weighted count of a number of occurrences of the URL component.

17. A system for adaptively filtering URL messages routed across a network, the URL messages rejected based on a set of rules, the system comprising:

a learning engine adapted to perform the steps of:

storing rejected URLs in a trie structure, wherein each node in the trie structure is associated with a URL component;

selecting a node according to a set of constraints; and

generating an exception rule for the selected node and its descendants; and

a filter configured to apply the exception rule to determine whether to allow the selected node and its descendants.

18. The system of claim 17, wherein each node associated with a URL component maintains a frequency, wherein the frequency is a function of a number of occurrences with which a URL component associated with a node was rejected with a rule and a number of occurrences with which descendants of the URL component were rejected with the rule.

19. The system of claim 18, wherein the set of constraints is selecting a node with a number of occurrences exceeding a threshold and having no children with a number of occurrences above the threshold.

20. The system of claim 18, wherein the set of constraints is selecting a node with a frequency exceeding a threshold.

21. The system of claim 19, wherein the threshold is a product of a total number of URL messages over a time interval and a percentage of the messages that should be allowed.

22. The system of claim 17, wherein the exception rule is generated by inferencing a scalar data type of the descendants of the selected node.

23. The system of claim 18, wherein the frequency is a direct count of the number of occurrences of the URL component.

24. The system of claim 18, wherein the frequency is a weighted count of the number of occurrences of the URL component.

25. A computer program product comprising:

a computer-readable medium having computer program code embodied therein for adaptively filtering URL messages routed across a network, the URL messages rejected based on a set of rules, the computer program code adapted to:

store rejected URLs in a trie structure, wherein each node in the trie structure is associated with a URL component,

select a node according to a set of constraints; and

generate an exception rule for the selected node and its descendants.

26. The computer program product of claim 25, wherein each node associated with a URL component maintains a frequency, wherein the frequency is a function of a number of occurrences with which a URL component associated with a node was rejected by a rule and a number of occurrences with which descendants of the URL component were rejected with the rule.

27. The computer program product of claim 26, wherein the set of constraints is selecting a node with a frequency exceeding a threshold and having no children with a frequency above the threshold.

28. The computer program product of claim 26, wherein the set of constraints is selecting a node with a frequency exceeding a threshold.

29. The computer program product of claim 25, wherein the computer program code is further adapted to apply the exception rule to determine whether to allow the selected node and its descendants to pass.

30. The computer program product of claim 27, wherein the threshold is a product of a total number of URL messages over a time interval and a percentage of the messages that should be allowed to pass.

31. The computer program product of claim 25, wherein the exception rule is generated by inferencing a scalar data type of the descendants of the selected node.

32. A computer-implemented method for adaptively filtering URL messages routed across a network, the URL messages rejected based on a set of rules, the method comprising:

storing rejected URLs in a trie structure, wherein each node in the trie structure is associated with a URL component;

selecting a node in the trie structure according to a set of constraints; and

generating an exception rule for the selected node and its descendants.

33. The method of claim 32, further comprising maintaining a frequency for each node associated with a URL component, wherein the frequency is a function of a number of occurrences with which a URL component associated with a node was rejected by a rule and a number of occurrences with which descendants of the URL component were rejected with the rule.

34. The method of claim 32, further comprising applying the exception rule to determine whether to allow the selected node and its descendants.

35. The method of claim 32, wherein the set of constraints is selecting a node with a number of occurrences exceeding a threshold.

36. The method of claim 32, wherein the set of constraints is selecting a node with a number of occurrences exceeding a threshold and having no children with a number of occurrences above the threshold.

37. The method of claim 35, wherein the threshold is a product of a total number of URL messages over a time interval and a percentage of the messages that should be allowed.

38. The method of claim 32, wherein the exception rule is generated by inferencing a scalar data type of the descendants of the selected URL component.

39. The method of claim 33, wherein the frequency is a direct count of the number of occurrences of the URL component associated with the selected node.

40. The method of claim 33, wherein the frequency is a weighted count of the number of occurrences of the URL component associated with the selected node.